Virtual Institute of Microbial Stress and Survival
Progress Report: April, 2003

I. Overview

This is the second of the monthly progress reports for the VIMSS Genomes To Life Project. Each core is expected to report on the progress of the technical and administrative actions from the previous month as well as forecast upcoming progress. I want to remind everyone how important to make sure everyone is communicating. The discussion boards (http://genomics.lbl.gov/~aparkin/discus) provide a forum for people to ask questions about direction of the project, priorities, and technical issues that can be read and answered by the entire group. I know email is often the most efficient means but it does privatize some of the important communications. Also, posting project data and information to BioFiles (https://tayma.lbl.gov/perl/biofiles) is EXTREMELY important. We are in the process of adding user help files to BioFiles – if you have user questions, please contact Keith Keller (tel: 510.495.2766 or email: kkeller@lbl.gov). This is the best metric I can give to the DOE leadership that we are making progress aside from the genomics websites. Please make us and yourselves visible by donating data and information to the website. Every PI should have at least one piece of substantive information in the database by next month.

II. Applied Environmental Microbiology Core

Lawrence Berkeley National Laboratory

SR-FTIR. We continue to modify the existing SR-FTIR spectromicroscopy apparatus to study *Desulfovibrio vulgaris* under anaerobic conditions. We are testing the modified system using reference samples. We identified that the most probable source of the problem was associated with the changes in the optical components of the apparatus. We expect to complete the apparatus for the live viewing of D. vulgaris on the SR-FTIR in May.

PLFA assays for Oxygen exposure. Humidified air was purged through 8 ml cultures of *D. vulgaris* for up to 22 hrs. Samples were removed at lyophilized in duplicate and analyzed for phospholipid fatty acid (PLFA) composition. Only minor changes in lipid composition were seen throughout between exposure times (see Figure 1 and Figure 2).

There was an increase in total lipid recovery after 5 minutes exposure. Further experiments will be performed to see if the variation in lipid recovery is due to oxygen exposure or to experimental variability. In this experimental run, overall lipid loss after 22 hrs was 6% (Figure 1).

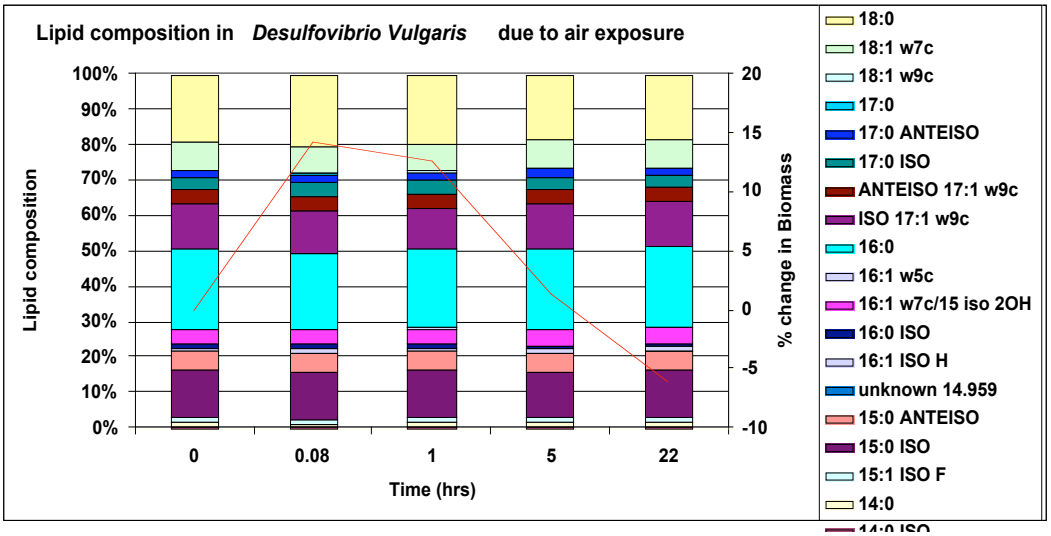


Figure 1. *D. vulgaris* PLFA composition after air exposure

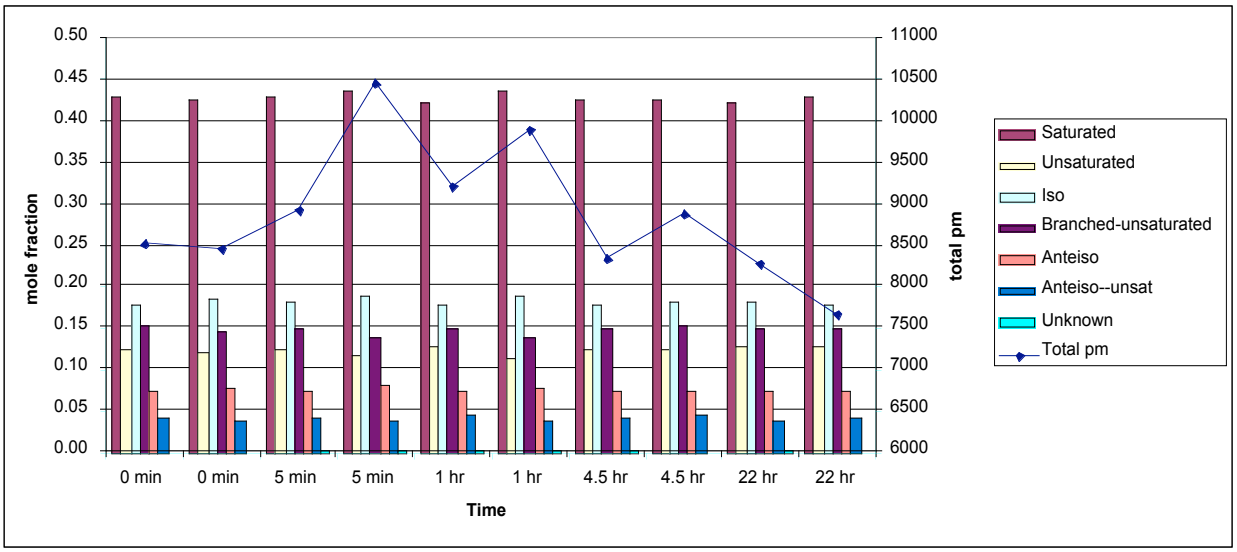
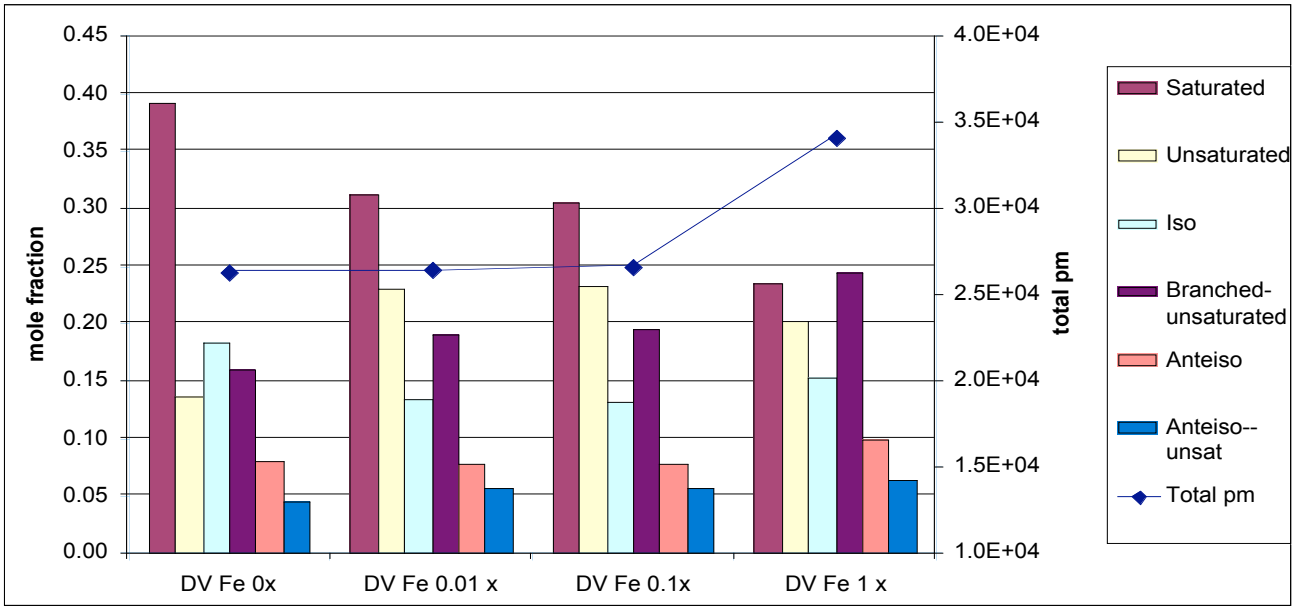


Figure 2. Same data as Figure 1, reorganized to show lipid guilds rather than individual lipids.

PLFA assays for Fe exposure. *D. vulgaris* cultures were grown in parallel cultures with Fe concentrations at full strength in Baars medium (1 x), 0.1 x full strength, 0.01 x full strength, and no Fe. Biomass, as measured by total pm lipids, was constant for the lower three Fe concentrations and increased by 20% in the full strength Fe medium.

The signature lipids changed with the Fe concentrations. The saturated lipids steadily declined and branched unsaturated lipids increased with increasing Fe concentrations.



MPN enrichments were set up for two subsamples of each sediment core from FB609-10-00 (background area, core depth 180-239in), FB060-01-00 (area 1, depth 180-228 in), FB052-01-00 (area 2, depth 245-300 in) and FB056-01-00 (area 3, depth 194-198 in) on B2 medium reformulated to increase its buffering capacity and supplemented with one of the following substrates: 1) lactate, 2) acetate, 3) propionate 4) pyruvate and 5) hydrogen with carbon dioxide. No sulfate reduction has yet been observed.

LS medium for SRB cultivation formulated by Judy Wall and co-workers was evaluated for growth of *D. vulgaris* Hildenborough ATCC29579 and *Desulfovibrio* sp. PT2. Both strains grown on lactate and sulfate demonstrated similar biomass yield which was comparable to that obtained on B2 medium. In order to insure comparability of growth and stress response studies conducted at the different participating labs, we will use the LS medium in our initial chemostat experiments.

We have initiated studies of two different *Desulfovibrio* species (*D. vulgaris* Hildenborough and *Desulfovibrio* sp. PT2) growing syntrophically with a hydrogenotrophic methanogen (*Methanococcus maripaludis*) in the absence of sulfate. These defined co-cultures offer an environmentally relevant growth state that could not be achieved in pure culture. Initial physiological studies now in progress will provide the foundation for future stress-response experiments. Although *M. maripaludis* is a marine isolate routinely grown on McC medium containing 22g/L of NaCl and 6.5g/L, initial studies completed in association with the John Leigh laboratory at the University of Washington revealed excellent growth of the methanogen on a low salt medium. Also, both strains of *Desulfovibrio* were shown to grow well on media supplemented with 10g/l NaCl.

We investigated several alternative protocols for DNA extraction from FRC site sediment. DNA yields were very low in all cases. Using an alternative approach based on washing of cells from sediment prior to disruption, we were able to isolate DNA of

sufficient quantity and quality for PCR amplification, but only with prior seeding of sediment with bacterial cells (*Methylobacterium extorquens* AM1) at 10^6 cells/ g of sediment.

Immediate future work

- Development of sulfate-reducing enrichments from deeper sediments from the FRC site.
- Continued isolation of SRB from sulfate-reducing enrichments
- Continued development and characterization methanogen and SRB co-cultures

Oak Ridge National Laboratory

Four sulfate-reducing cultures were initiated from freezer stocks and shipped to Sergey Stolyar at University of Washington: three from sediments in Area 3 and one culture from groundwater in Area 2. We have also cultured (4 l) one enrichment from FW-109 sediment (10-2 dilution) and extracted nucleic acids. We are in the process of constructing a SSU rRNA clonal library as well as attempting the construction of a large insert fosmid library.

Diversa

- 5 cored soil samples have been received from LBNL.
- Small insert DNA has been extracted from 4 samples (top and bottom). Large insert DNA extractions have not resulted in visible concentrations of DNA. These samples are being amplified with Phi29 polymerase.

ample ID	Core Segment	Core Depth	Date Received	Shipped By	Comments	Small Insert DNA	Large In DNA
WB302-03	? (~10" long)	?	?	LBNL	Background	4ng/ml top 3ng/ml bottom 3/27/2003	Nothi
B609-01-00	24-43" long	180-234"	3/5/2003	LBNL	Background, Anaerobic	5ng/ml&4ng/ml top 4ng/ml x2 bottom	Nood Nood
B060-01-27	27-46" long	180-228 "	4/10/2003	LBNL	<25ppm U238 Anaerobic	5ng/ml top 5 ng/ml bottom 4/24/2003	Noodl Noodl
B052-01-12	12-36" long	245-300"	4/10/2003	LBNL	<25ppm U238 Anaerobic	5ng/ml top 5 ng/ml bottom 4/24/2003	Noodl Noodl
B056-01-34	34-58" long	144-198"	4/10/2003	LBNL	<25ppm U238		

- *E. coli* genomic DNA was prepared and amplified using Phi29 DNA polymerase. This amplified DNA as well as the original genomic DNA were hybridized to an Affymetrix GeneChip *E. coli* genome array. Results indicate that overall coverage of the *E. coli* genome in the amplified sample is comparable to un-amplified control DNA. There were 39 genes that showed preferential amplification with the Phi29 (10 fold difference from unamplified.)

- The amplified DNA was examined on a pulsed field gel, and shown to have sizes up to ~50 kb.

Issues:

- Low DNA yields from the contaminated soil cores may require amplification before library construction.

- DNA amplification with Phi29 polymerase results in amplification of non-specific products in the negative control samples, which may interfere with efficiency of library construction.

Actions:

- Large and small insert DNA extractions will be completed for the remaining soil sample by 5/5/03.

- Phi29 polymerase will be used to amplify the small and large insert DNA extracted from the LBNL soil samples, so that enough DNA is obtained for library construction.

- Experiments are being conducted to increase the size of Phi29 amplified DNA.

- *S. diversa* genomic DNA, which is GC rich, will be amplified with Phi29 and hybridized to an Affymetric GeneChip to look at genome coverage and product size.

- The Phi29 amplified *E. coli* genomic DNA has been digested with library construction enzymes. This sample will be used to construct a large insert library, to determine if workable titers are obtained from amplified DNA.

- Work is ongoing to optimize the large insert FACS Biopanning protocol in gel microdroplets. Recovery of DNA directly from hybridized GMDs has been shown to be feasible.

- A temperature gradient was conducted with the Phi29 polymerase reaction. At the normal temperature of 30^{o}C and at 25^{o}C, DNA yields were good, but there was non-specific product in the negative control. When temperatures were increased to 33^{o}C and 37^{o}C, the yield decreased slightly, but there was no visible product in

the negative control. The higher temperatures may be necessary to obtain specific DNA amplification, which will increase library efficiency.

III. Functional Genomics Core

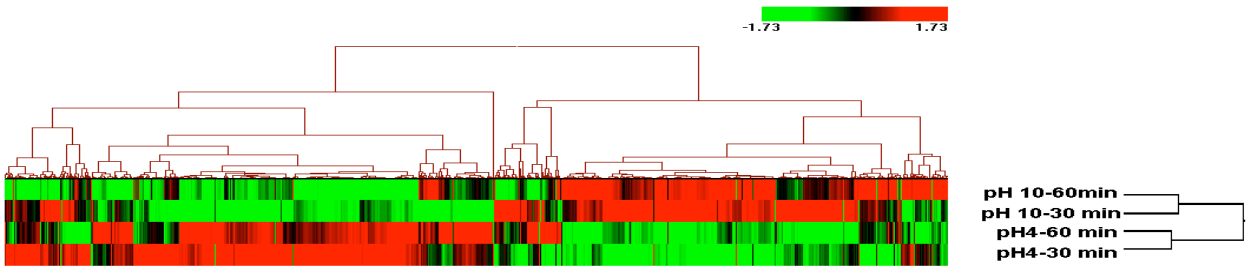
Transcriptomics (ORNL)

Objectives

- To perform initial microarray expression profiling studies for the model bacterium *S. oneidensis* MR-1 to establish a baseline response to various environmental stresses.

Progress since last report

- Completed detailed analyses of microarray expression data from pH stress experiments comparing growth of MR-1 under neutral pH (pH 7) to growth under pH 4 or pH 10. In this study, we have examined the transcriptome of *S. oneidensis* MR-1 during its response to both low- and high-pH stress using DNA microarrays covering ~99% of the total predicted protein-encoding open reading frames in this bacterium. Total RNA was isolated from cells, which were grown to mid-exponential phase, and subsequently exposed to either acidic (pH 4) or basic (pH 10) conditions for 30 and 60 minutes. Microarray data were further analyzed using pairwise average-linkage clustering (Figure 1). In addition, genes showing differential expression in response to pH stress were grouped according into functional categories (Figure 2). A subset of genes differentially expressed in response to pH stress has been selected for further confirmation using real-time PCR. The writing of a manuscript describing this work is currently underway. A poster (H-003) describing this work will be presented at the 2003 ASM General Meeting.



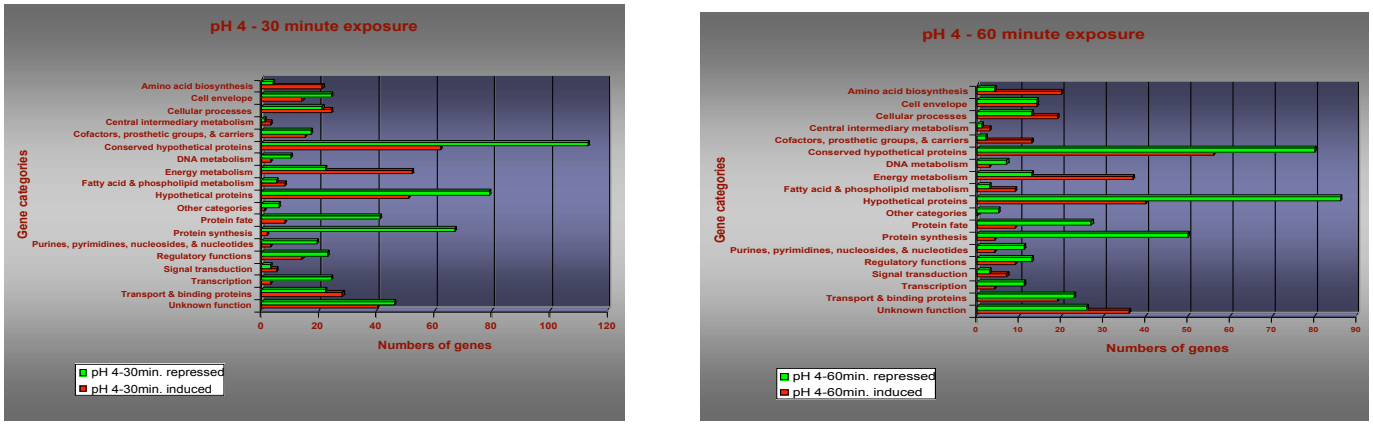


Figure 1. Hierarchical Cluster Analysis of pH Stress Microarray Data

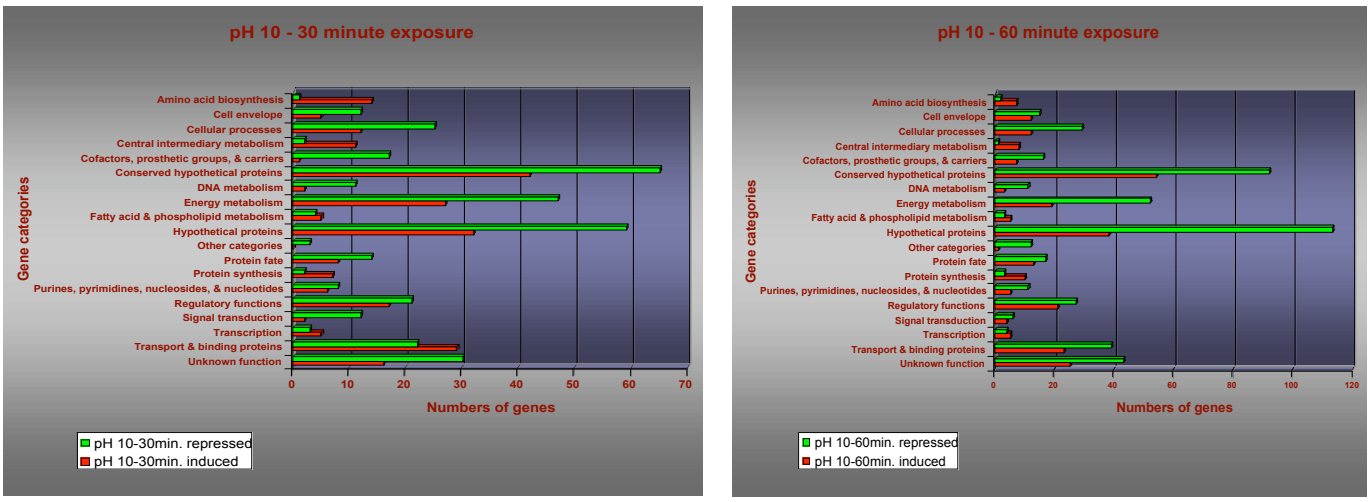


Figure 2. Functionally grouped genes showing a 2-fold or greater difference in mRNA abundance in response to pH 4 and pH 10.

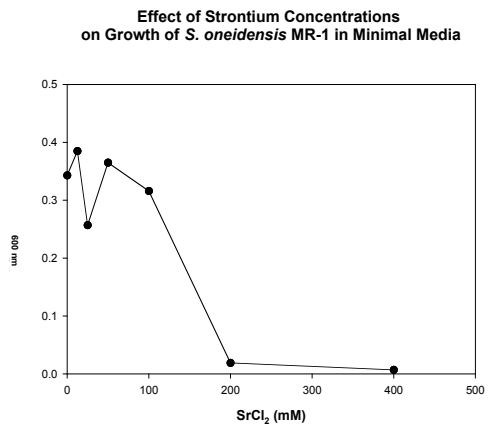
- Microarray experiments analyzing the MR-1 transcriptome under heat shock conditions have been completed and the data analyzed using Genespring. Transcriptome profiles under nonshock conditions (30°C) were compared to those obtained under shock conditions (42°C) for 5, 15, and 25 min. Homologues for the classical sigma 32-regulated genes are showing increased expression in response to heat shock at 42°C (see table below). A poster describing this work will be presented at the 2003 ASM General Meeting.

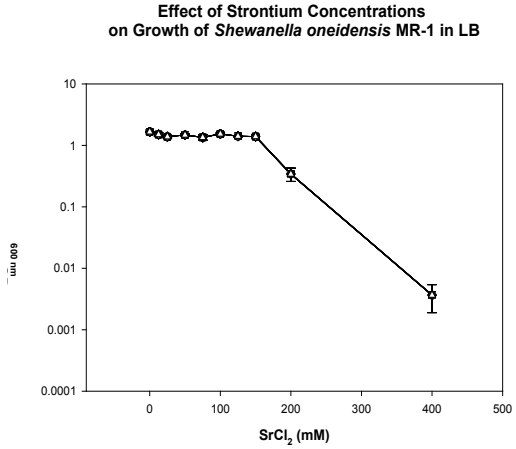
Induction of known heat shock genes

Gene	Microarray results (fold induction at time indicated)		
	5min	15min	25min
dnaK	4.7	3.3	2.3
dnaJ	3.8	4.6	3.4
grpE	5.9	5.2	1.8
groES	3.6	3.2	3.8
groEL	2.4	2.3	2.3
clpB	9.0	6.8	4.5
htpG	5.9	4.6	4.3
ftsH	3.3	2.9	1.4
lon	3.8	3.2	2.2
ibpA	No data	No data	No data
rpoH	3.6	3.6	2.3
rpoE	1.6	1.4	1.2

- Raw and statistically analyzed pH stress response data have been uploaded into Biofiles. Raw heat stress response data have been uploaded into Biofiles.

- Preliminary growth studies looking at the response of wt MR-1 to strontium toxicity have been completed. These experiments indicate that MR-1 growth is inhibited strongly at strontium chloride concentrations of about 200 mM in both minimal and rich media.





- Microarray experiments analyzing the effect of H_2O_2-induced oxidative stress on global gene expression patterns in wt *S. oneideinsis* have been completed. Preliminary analysis of the microarray data suggest that the major hydrogen peroxide-responsive genes include cellular detoxification genes, such as *ahp*C, *ahp*F, and *kat*B, DNA-binding protein gene *dps*, as well as genes involved in the TonB transport systems. Comparison of the time- and dose-dependent gene expression kinetics revealed the transient nature of some of the H_2O_2-induced

genes and the selectivity of different H_2O_2 stress conditions on the regulation of gene expression. A poster describing this work will be presented at the 2003 ASM General Meeting.

Future work

- Complete pH stress manuscript. Design more detailed time-series experiments.
- Complete detailed data analysis for oxidative stress microarray experiments.
- Begin microarray expression profiling of MR-1 cells exposed to toxic levels of certain metals (e.g., strontium).
- Phenotype characterization and gene expression analysis of various mutant strains (e.g., ΔompR, ΔenvZ, ΔoxyR, Δdps).

Proteomics (Diversa)

Objectives

- Proceeded to the completion of *Desulfovibrio vulgaris* proteome analysis by working on the last fraction -SDS fraction.
- Develop and evaluate different methodologies for differential quantitative proteomics analysis using chemical modifications. With the intensity based quantification methodology (see last month report), we are trying to establish a set comprehensive tool to quantify the change of protein expression upon cell stress.

Progress since last report

- We finished the remaining analysis of the *Desulfovibrio vulgaris* proteome (data posted on Biofiles).
- Upon the completion of sample preparation optimization and software development, our activities moved towards the method development for quantitative proteomics.
- Sample analysis: SDS fraction of *Desulfovibrio vulgaris* has been analyzed by Diversa's platform and data analysis is about to proceed.

Future work

- Method development and evaluation in chemical/isotope labeling procedures for differential quantitative analysis.
    - Esterification methods: it target C-terminal and carboxyl group of peptides. We further optimized the protocols. The yield has been significantly improved on standard peptides. We are going to apply this on peptide mixtures of higher complexity, e.g. standard protein digests and whole cell lysate digests.

    - Sulfohydral labeling reagents: we have designed several new compounds to label the reduced cysteine side chain to introduce a differential mass.

The starting material has been purchased and the synthesis of these chemical is about to start.

- ICAT reagent: it is the commercial available reagent for quantitative proteomics using isotope labeling. We have purchased the reagent to compare our own developing method.

Protein complexes (Sandia)

Objectives

- Identify target proteins to be tagged for protein complex identification.
- Use proteomics techniques like ICAT and others to identify proteins that are overexpressed under stress conditions of heat, O2, pH and UV radiation.

Progress since last report

D. vulgaris:

- Diagnostic PCR runs for the nitrite reductase gene (*nfr*H) from *D. vulgaris* (ATCC and U. Missouri cultures) were successful. Extremely faint bands of cytochrome c3 from *D. desulfuricans* (possibly contaminating strain) were observed (after 35 cycles) for both ATCC and U. Missouri cultures - perhaps due to non specific binding – at least 10 bp of the corresponding forward primer of cytochrome c3 are present in multiple locations within the *D. vulgaris* genome.
- ICAT experiments were performed (with LBNL/UCB) for *D. vulgaris* proteome using mcapillary LC-MS-MS. The no-heat-shock control condition (light isotope label) was compared to a 30 min heat shock experiment (heavy isotope label).
- ProID software was used to identify labeled peptides using in-silico digested proteome of *D. vulgaris* as the interrogator database. A total of 219 proteins were identified. Out of these 7 are classified as stress related proteins based on currently available annotation.
- We are currently working on the ProICAT software to quantify differences in expression levels between the two sample populations.
- Also efforts have been directed towards a more sensitive analysis using nano-LC columns for the same sample population. Protocols for in-gel digests (SDS-PAGE) and subsequent identification through LC-MS-MS and database searches for standard proteins (a-amylase from *Bacillus amyloliqueficans*, b-galactosidase from *E. coli* and bovine serum albumin) were successfully established. These protocols will be used in near-future studies for the identification of components of protein complexes from *D. vulgaris*.

E. coli

- UV Radiation experiments on *E. coli* K12 (pET20) were repeated.

- *E. coli* cells (100 ml) in mid-exponential phase (OD600 ~0.35) were placed on a continuous shaker in a laminar-flow bio-hood equipped with a UV-C source.
- Cells (10 ml) were withdrawn every 30 s and placed on ice for 5 min and spun down at 7000 rpm for 20 min at 4C.

Future work

- Analyze *D. vulgaris* and *E.coli* samples using nanoLC-MS-MS, ICAT and other techniques.
- Magnetic beads for immunoprecipitation experiments have been ordered from Dynal Biotech.

Metabolomics and Proteomics (UCB, LBNL)

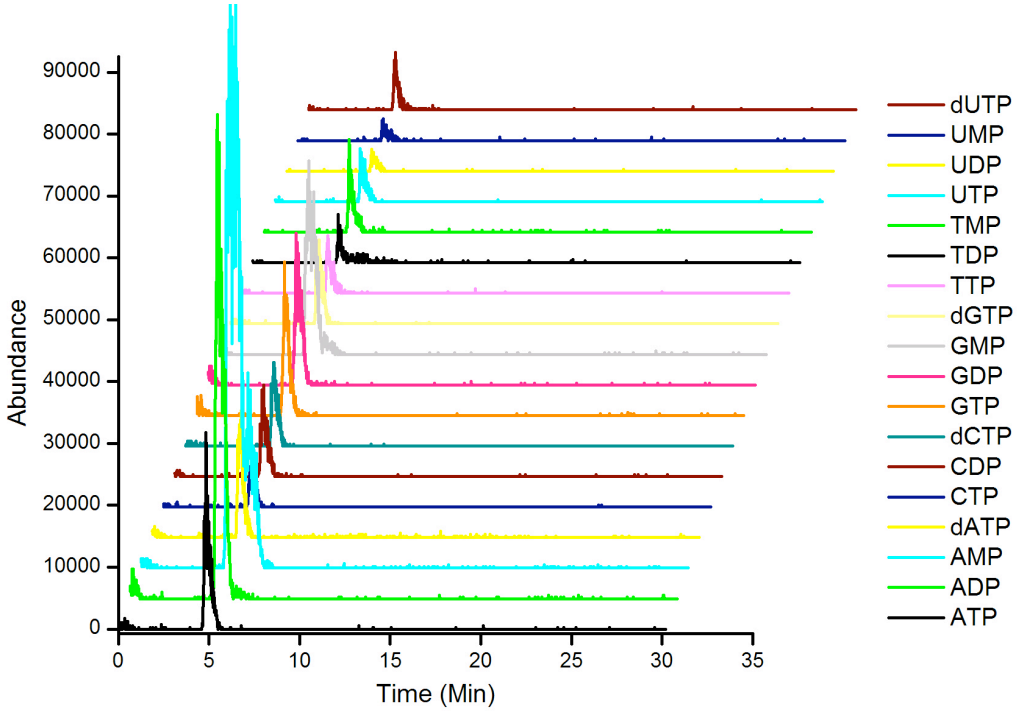
Objectives

- Development rapid and sensitive methods for identification and quantification of nucleotides, carbohydrates and their respective derivatives.
- Run preliminary ICAT experiments on heat-shocked samples of *D. vulgaris* (with Sandia) and sporulation of *Bacillus* to establish procedure in the lab.

Progress since last report

Metabolomics:

- 18 nucleosides mono/bi/tri phosphate and deoxy-nucleotides standards were successfully detected using different scan modes, including EMS scan, product ion scan and MRM scan.
- A reverse phase column based separation method was evaluated for the separation of these compounds. At pH 6.8 and pH 5 in 10 mM NH_4OAc buffer, none of them is retained but the signals are seriously suppressed at pH 5.
- A high signal/noise ratio can be obtained without using any buffers, but the signal detections are not consistent. A low concentration of NH_4OAc is advantageous for the analysis of trace amount of samples.
- Preliminary analysis of carbohydrates and carbohydrate derivatives was initiated. The MS-MS conditions required to analyze simple neutral (glucose) and sugar alcohols (sorbitol) were determined. Positive ionization (using the metal adducts) and negative ionization were tested.



Example of detection of 18 nucleosides/nucleotides phosphate derivatives by multiple reaction monitoring (MRM). LC conditions are following: 10 minute washing with buffer A (water only) and then 20 minutes gradient to 20% ACN. Signals are scanned by a tandem mass spectrometer in EMS scan mode (data not shown) and MRM scan mode. A C-18 column was used in the development.

ICAT:

D. vulgaris

- A strategy to study global stress response in *D. vulgaris* is being developed. In bacteria, responses to stress and changes in the environment are often regulated via two component systems comprised of Histidine Sensor Kinases (HSK) and Response Regulators (RR). Using the latest annotated FASTA file of the sequenced *D.vulgaris* genome, all ORFs annotated as HSK were analyzed *in silico*. Around 30 ORF were annotated as HSK(s) or related protein(s). Of these 13 had the characteristic conserved H, N, D/F, anf G boxes (See Figure). One interesting observation is that 9 of these are "hybrid", and also contain a receiver domain at their N terminal end. Primers are being designed for each of these HSK(s) to generate a library of knock-out mutants in *D. vulgaris*. A first model candidate for this study was based on the results of the heat shock experiment with *D. vulgaris* (Sandia National Laboratories). The ICAT data set generated (see Biofiles) identified *orf* 00061, a HSK. The 2001bp *orf* 00061 was amplified and cloned into the pET30a(+) vector as a *Bam*HI *Eco*RI fragment. The MCS of pET30a(+) will allow subcloning of a fragment containing *orf* 00061 into pLO1 (obtained from Judy Wall). pLO1 will be used to generate the knock out mutant of *orf* 00061 in *D. vulgaris*.

B. subtilis 168

- Protein fractions were collected from *B. subtilis* 168, after the initiation of sporulation via a nutrition deprivation stress at time points t0 and t30 mins. The two proteomes were ICAT labeled and subjected to tandem nanoLC-MS-ESI analysis. Preliminary data, comprised of a list of target proteins identified by the ProID software will be uploaded to Biofiles shortly.
- Comparison of the two proteomes using the ProICAT software led to the identification of proteins that demonstrated differences 30 mins after initiation of sporulation. Scrutiny of this data set is in progress. Of those examined proteins that show the same trend as that of the transcriptome data (see Microarray data, Biofiles) are highlighted in red while those that show an opposite trend are blue (see Biofiles).
- Primers have also been obtained for *orf* 5707, a hybrid histidine sensor kinase, containing not only a receiver domain but also a HPt domain (homologous to ArcB) at its N terminus.

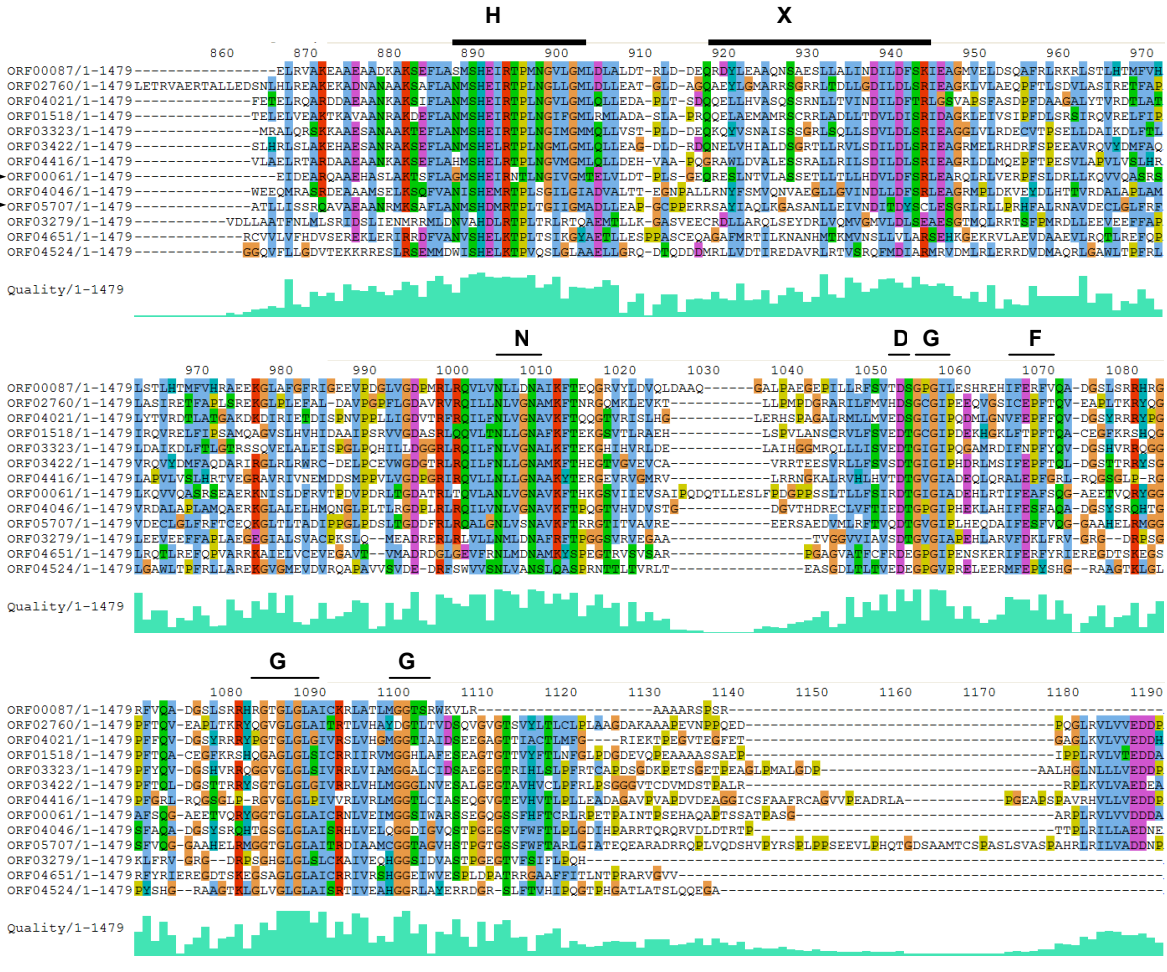


Figure. Conserved regions of Histidine Sensor Kinases in *Desulfovibrio*

Future work

- Hydrophilic interaction and ion-exchange HPLC methods will be evaluated for the separation of nucleotides and derivatives. Due to the compatibility of buffers and organic solvents with electrospray ionization, we are still going to focus on NH_4OAc buffer and ACN/methanol, but alternative buffers will be evaluated.

- Continue on the analysis of carbohydrates and derivatives

- Work toward completing an ICAT experiment for oxygen stress response (with Hazen Lab).

- The differences in the data set generated by the ProID and ProICAT software have not yet been reconciled. Scrutiny of the two data sets is in still in progress.

IV. Computational Core

In April, we achieved the first of our long-term project goals – to produce cis-regulatory element predictions for Desulfovibrio vulgaris. These include the sequence of the predicted cis-regulatory element, a list of operons downstream from each motif, as well as any available annotations for the genes comprising each downstream operon. These preliminary results have been uploaded to Biofiles, and we invite requests for more in-depth analysis of specific interactions of interest to other team members.

We also began collecting data for an in-house comprehensive microbial gene expression database. The initial source of information for our database comes from a list of about 100 publications from 67 labs. Thus far, we have obtained data from about half of these labs via public websites and email correspondence. Currently, the database includes microarray data from nearly 700 experimental conditions in a diverse set of bacteria.

We hired and trained a new bioinformatics specialist in April, and with her help, made progress in migrating some key applications to our new comparative genomics database, including our operon prediction method. We hope to prepare a manuscript describing this method for publication in the near future. Due to unexpected delays in the shipping of several components, we are only about halfway done installing the computer infrastructure we purchased.

During the next months, we plan to complete cis-regulatory element prediction for two other target organisms, Geobacter metallireducens and Shewanella oneidensis. Also, we plan to present a comparison of gene regulation in these related organisms. We will also begin to improve these predictions by using critica-based gene models, and by incorporating our own operon predictions. After the installation of our computer cluster is complete, we will begin to expand the number of genomes in our comparative microbial genomics database to include all those currently sequenced to draft quality. Finally, we expect to produce a prototype for a comparative genome browser within the next month.

Biopathways Graph Data Manager (LBNL—Olken)

Objectives

1) Commence design of context annotations for biopathways graphs.
2) Specify graph characterization queries to be supported.

Progress since last report

Data Modeling

- We added k-core queries, and graph characterization queries (per Denise Wolf) to our list of query types to be supported. See web page for details.

- We began work on context representation for bioprocesses (reactions). Contextual annotations are used to specify the circumstances under which reactions occur: Contexts may be variously: phylogenetic (which organism), anatomical, genotypic, developmental, pathological, or environmental (pH, aerobic vs. anaerobic, freshwater vs. salt water, temperature range). We view contexts as Boolean predicates, e.g., described in a description logic which includes inequality predicates. It should also be possible to name contexts and refere to them by name. See project web page for details.

- We began work on the evaluating the possible use of RDF as our graph data model. RDF is the Reseource Description Framework, developed by the World Wide Web Consortium. (W3C). It is a metadata format based on a simple graph data model. It is being used as the basis for the Semantic Web Activity – an effort to facilitate machine processable documents on the web with formally specified semantics. The attraction of RDF is that there is an existing community developing tools for RDF.

Papers

- Frank Olken and Doron Rotem, "Workflow Execution History Data Management: A Framework", was submitted to International Conference on Web Services, Las Vegas, NV, July 2003. This paper discusses a graph data model for workflow execution history data. It grew out of our graph data manager work and is relevant to LIMS (Laboratory Information Management Systems).

- Work continued on the workshop report and the special issue of OMICS both from the workshop Olken helped to organize at National Library of Medicine in Feb. on Database Management for Life Sciences Research.

Web Page

Our project web page, http://www.lbl.gov/~olken/graphdm/graphdm.htm, was revised in April to reflect changes in query types, graph characterization queries, and context annotations.

Infrastructure

We took delivery of two Linux workstations and got them up and running using the Mandrake 9.1 distribution.

Future Work

1) Elaborate design of context annotations and queries for biopathways (sub)graphs.

2) Commence design of visual graph-based query language.

3) Complete workshop report from NLM Workshop on Data Management for Life Sciences Research.